Stochastic approach to molecular interactions and computational theory of metabolic and genetic regulations

H. Kimura, H. Okano and R.J. Tanaka*

Bio-Mimetic Control Research Center, RIKEN,
Shimo-shidami, Moriyama-ku, Nagoya 463-0003, Japan

Abstract

Binding and unbinding of ligands to specific sites of a macromolecule are one of the most elementary molecular interactions inside the cell that embody the computational processes of biological regulations. The interaction between transcription factors and the operators of genes and that between ligands and binding sites of allosteric enzymes are typical examples of such molecular interactions. In order to obtain the general mathematical framework of biological regulations, we formulate these interactions as finite Markov processes and establish a computational theory of regulatory activities of macromolecules based mainly on graphical analysis of their state transition diagrams. The contribution is summarized as follows:
(1) Stochastic interpretation of Michaelis-Menten equation is given.
(2) Notion of *probability flow* is introduced in relation to detailed balance.
(3) A stochastic analogy of *Wegscheider condition* is given in relation to loops in the state transition diagram.
(4) A simple graphical method of computing the regulatory activity in terms of ligands' concentrations is obtained for Wegscheider cases.
(5) A general comutational scheme of the stationary probability distribution is obtained in terms of probability flows.
keywords: Biological regulation, molecular interaction, Markov process, stationary probability distribution, probability flow

1 Introduction

Control is a ubiquitous built-in mechanism supporting a variety of functions of living organisms. The advent of cybernetics in the late 1940's established a close link between control in living organisms and that of man-made artifact. Notions like feedback, feedforward, robustness, stability, and noise, which are fundamental in control engineering turned out to be relevant to the control mechanisms of living organisms. Rapid progress in molecular biology in the last fifty years opened up a new world of intracellular control of biochemical processes including those of metabolism and gene expression. Studying control mechanisms became an important issue of biology. Monod correctly called this field *microscopic cybernetics* [31], while others called it *regulatory biology* [50]. The continuing endeavor to clarify the material basis of regulations reveals vast complexity of regulatory networks associated with huge variety of material links inside a cell, which in turn motivated considerable interest in the quantitative analysis of regulation mechanisms through mathematical modeling and simulation. The modeling/simulation paradigm has been used to reveal the underlying structural properties of complex regulatory networks. Existence of cellular switches [1][8][14][23][30][34][37], evaluation of molecular fluctuations [16][21][28][43], analysis of biological oscilations [5][7][17], consideration of network robustness [25][51], theoretical demonstration of cellular behaviors [11][12][39], are examples of its contributions. Excellent reviews of these achievements are found in [41][44][45]. Recently, we proposed the notion of compound control [42] which captures a salient characteristic feature of biological control. Compound control is a biological way of realizing proper input/output relationship of regulatory dynamics to cope with diverse, complicated and unpredictable environmental changes.

It is paradoxical that the theoretical and/or physical basis of the modeling/simulation paradigm has not been well exploited, in spite of the remarkable success it has attained. For instance, in many papers on intracellular regulations, the traditional

*Corresponding author. Tel.: +81(52)736-5861, Fax: +81(52)736-5862, Email: reiko@bmc.riken.jp

Michaelis-Menten-Hill equation in enzymology is extensively used for describing molecular interactions, e.g., the action of repressors in transcription regulation, or the effect of feedback [5][14][27][30][34][48], without sound justification.

The transcription rate is regulated through binding of transcription factors in some domains of DNA, the so-called *cis*-regulatory elements [6], that are able to affect transcription initiation. The affinity of *cis*-regulatory elements to the transcription factors determines the overall transcription rate of the gene. They are correlated with each other and collectively exhibit synergistic effects through which a complex computation is performed to achieve adequate transcription control [26]. These factors are represented in the context of a thermodynamical equilibrium energy distribution suggesting to plausible models of transcription regulation [3][4][35][47]. This paper aims to establish a unified framework with solid mathematical grounding and physical plausibility that is able to capture various bio-physical attributes of intracellular regulations.

A significant portion of genes encode enzymes which are involved in metabolic control, another important form of intracellular regulation. Metabolic control is a relatively old subject of bio-chemistry, where allosteric enzymes are the major agents of control. Allosteric enzyme is a macro-molecule that has several binding sites which can bind substrate and ligands. The binding pattern of ligands changes the conformations of the enzyme, through which the activity of the enzyme for the target reaction is controlled. Quantification of the activity changes of allosteric enzyme has been a central issue in biochemistry since the classical Monod-Wyman-Changeux model was first proposed [32][49]. Since then, various theoretical and experimental methods have been used to predict the conformation changes due to substrate binding and modification by ligands [3][9][13][15][18][24][29][36][40][46][52].

The regulatory mechanism of allosteric enzymes is in some sense similar to that of operons, in spite of the obvious functional and structural differences between them. Both of them are controlled by interactions between binding sites (catalytic sites for metabolic cases and *cis*-regulatory elements for genetic cases) and binding factors (ligands and substrate in the metabolic case and transcription factors in the genetic case). In the metabolic case, interactions take place between proteins, while in the genetic case they take place between protein and DNA. It is interesting that although the allosteric protein is a player in both regulations, it plays totally different roles. In the metabolic control, it is controlled by ligands, while in the genetic control it acts as transcription factors to control genes.

The protein/protein interactions of metabolic regulation and protein/DNA interactions in transcription control are weak and essentially reversible, and they share many properties. In fact, equilibrium statistical mechanics is used to quantify both metabolic regulation [15] and genetic regulation [4][35][47]. If we view the binding/unbinding processes taking place between binding sites and binding factors in both cases as a common stochastic process, it suggests a common theoretical framework to deal with metabolic and genetic regulations in a unified way.

Along this line, we propose a common theoretical framework for describing biological regulators of metabolic and genetic regulations to facillitate quantitative descriptions of intracellular biochemical processes. The core of our approach is to describe protein/protein and protein/DNA interactions with a common Markov process focusing on the behavior of a single molecule rather than a population of molecules. This viewpoint enables us to describe phenomena with a finite state Markov process, which greatly simplifies the argument by avoiding the difficulties associated with infinite dimensionality [33]. All the quantification features associated with various complicated physical and biological phenomena can then be condensed into transition probabilities. We introduce a new notion of probability flow that offers a new insight into the stationary distribution of the Markov process. The stationary equation is regarded as representing the conservation of probability flows at each node. It is closely related to the Wegscheider condition derived a century ago. Based on the notion of probability flow, we derive a representation of the stationary probability distribution, which leads to a unified quantitative form for the general biological regulator. Our method of computing the stationary probability distribution dramatically simplifies the classical King and Altman method [24], which is used for computing the stationary distribution of chemical processes. It is expected to yield a new computational tool for operon regulation that can deal with the increasing complexity of operons [38].

Metabolic and genetic controls have been investigated relatively independently because of their apparent large differences in modality. Therefore, there have been few studies dealing with systems that combine metabolic and genetic regulations [52]. The unified framework described in this paper will enhance our understanding of intracellular biochemical processes.

In the next section, we formulate a biological regulator in an abstract way as a Markov process. The abstract biological regulator is defined through the stationary solution of the Master Equation. In Section 3, we introduce the notion of probability flow and explain its relevance to the stationary distribution. Sections 4 and 5 deal with the single loop case and introduce the Wegscheider condition, which guarantees that the probability flow vanishes. Section 6 generalizes the results of the preceding section to multi-loop cases. Section 7 discusses the meaning of transition probabilities to deduce biologically meaningful representations of biological regulations.

2 Characterization of biological regulations as a finite-state Markov process

The main stage of genetic control is transcription regulation whose fundamental mechanism is to change the transcription rate of the operon via binding/unbinding of transcription factors to and from their *cis*-regulatory elements. The main agent of metabolic control is the allosteric enzyme that determine the rate of the target chemical reaction via binding/unbinding of substrate and ligands. The fundamental common feature of both regulations lies in the molecular interaction between sites of the macro-molecule and molecular binding factors that work through binding/unbinding processes (see Table 1).

These processes are obviously random subject to certain thermodynamical constraints. Our idea is to establish a mathematical framework to describe such regulatory mechanisms in a unified way. We now formulate the biological regulation in a somewhat abstract way. We assume that the macromolecule, which is the main agent of the regulation, has n binding sites (BS), each of which can bind some of m ligands with different affinities that depend on the binding patterns of sites. Instead of the ligand, we use the word *binding factors* (BF) to emphasize the bilateral symmetry of sites to be bound and factors to bind. Each BS can bind only one BF at each time. The binding sites are denoted by $b_1, b_2, \cdots, b_n$, while the binding factors by $U_1, U_2, \cdots, U_m$. The state of the regulator is denoted by n-tuples of m alphabet $U_1, U_2, \cdots, U_m$ denoting the BFs being bound and ϕ which represents the empty (unoccupied) site. If the regulator has three BSs, for example, $S = (U_1 \phi U_2)$ denotes the state with b_1 and b_3 being bound by U_1 and U_2, respectively, and with b_2 empty. Let N be the total number of non-empty states, which is obviously finite. Since we have the empty state $S_0 = (\phi\phi\cdots\phi)$, the total number of states is $N+1$.

The state of the regulator changes as its binding pattern changes, i.e., as BFs bind to and dissociate from BSs. These are clearly stochastic phenomena and are legitimately described as stochastic process. Let the transition probability from the state S_j to S_i during an infinitesimal time duration Δt be denoted by $q_{ij}\Delta t$. All the quantitative features of the complex biochemical process associated with protein/protein interactions and protein/DNA interactions including origomerization [36][49], conformational change [13][35] and target localization [9] can be adequately represented in terms of q_{ij}. For the sake of clarity, we assume that binding or unbinding of only one BF to and from a BS occurs during the infinitesimal time duration Δt. Thus, the transition from $(U_1\phi U_2)$ to $(U_1\phi U_3)$ is regarded as consequtive transitions from $(U_1\phi U_2)$ to $(U_1\phi\phi)$ and from $(U_1\phi\phi)$ to $(U_1\phi U_3)$. The transition probability for unit time q_{ij} represents the rate of reaction from S_j to S_i, and sometimes it is regarded as identical to the rate constant of the corresponding reaction. We define $\mathbf{A}_i$ to be the set of states that are accessible from S_i during the infinitesimal time duration Δt. We call it the *adjacent set* of S_i.

Let $p_i(t)$ be the probability that the regulator is in a state S_i at time t. Then, the stochastic time-evolution of the regulator is described by the *chemical master equation* (CME),

$$\frac{dp_i}{dt} = \sum_{j\in\mathbf{A_i}} q_{ij}p_j - \sum_{j\in\mathbf{A_i}} q_{ji}p_i. \quad i = 0, 1, \cdots, N. \tag{1}$$

The notation $\sum_{j\in\mathbf{A}_i}$ denotes the sum of all j such that $S_j \in \mathbf{A}_i$. The first term of the right-hand side of (1) represents the net probability coming from adjacent states to S_i while the second term the net probability leaving S_i to adjacent states. The CME (1) is written in the matrix form as

$$\frac{dp(t)}{dt} = Qp(t) \tag{2}$$

where $p(t)$ denotes the $(N+1)$-dimensional vector whose $(i+1)$-th component is $p_i(t)$. The matrix Q in (2) is called a *transition matrix*.

A finite Markov process can be described by the state-transition diagram. In Figure 1 shows examples of state-transition diagrams. The most salient feature of the Markov process for our purpose is the reflection property that $q_{ij} \neq 0$ always implies $q_{ji} \neq 0$ because of the reversibility of the molecular interactions we are interested in this paper. In

Table 1: Comparison of genetic and metabolic regulations.

	Agent	Site	Factor	Control Parameters
Genetic	Operons	*Cis*-regulatory elements	Transcription factors	Transcription rate
Metabolic	Allosteric enzymes	Catalytic sites	Ligands and substrates	Reaction rate

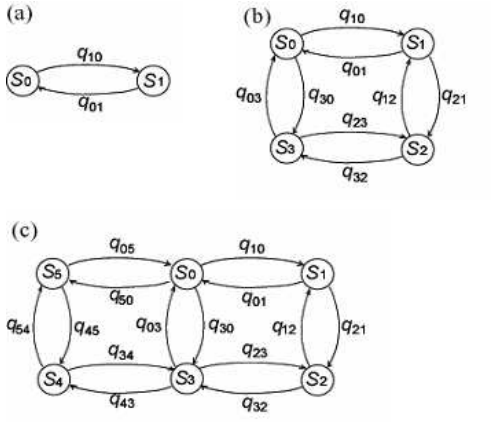

Figure 1: State transition diagram of (a) Example 1, (b) Example 2,(c) Example 3.

terms of the adjacent set $\mathbf{A}_i$, $S_j \in \mathbf{A}_i$ always implies $S_i \in \mathbf{A}_j$.

It is known [46] that the solution of (2) converges to a unique equilibrium stationary solution p that satisfies

$$Qp = 0, \tag{3}$$

or equivalently,

$$\sum_{j\in\mathbf{A}_i} q_{ij}p_j - \sum_{j\in\mathbf{A}_i} q_{ji}p_i = 0. \tag{4}$$

Equaiton (3) or (4) is called the *stationary state equation* (SSE). Usually, the convergence is faster than the chemical reactions controlled by the regulator. Therefore, we can assume that the regulator is always in a stationary distribution satisfying (3), together with the normalization constraint,

$$\sum_{i=0}^{N} p_i = 1. \tag{5}$$

Let γ_i be the activity of the regulator at state S_i. The overall activity γ of the regulator is then defined as the average activity with respect to the stationary distribution p, i.e.,

$$\gamma = \sum_{i=0}^{N} \gamma_i p_i. \tag{6}$$

The examples that follow show that the above definition of the regulation activity adequately describes both genetic and metabolic regulations.

Example 1. (Stochastic version of Michaelis-Menten equation)
Consider a molecule of an enzyme E that catalyses the reaction of producing a product P from a substrate S, i.e.,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \overset{k_2}{\to} E + P. \tag{7}$$

An enzyme molecule can be regarded as a biological regulator with one BS and one BF (the substrate). The enzyme has the two state S_0 and S_1, the empty state and the state occupied by a substrate molecule, respectively. The transition diagram is shown in Fig. 1(a). The transition matrix Q in this case is written as

$$Q = \begin{bmatrix} -q_{10} & q_{01} \\ q_{10} & -q_{01} \end{bmatrix}, \tag{8}$$

and the stationary distribution is given by

$$p_0 = \frac{1}{1+\frac{q_{10}}{q_{01}}}, \quad p_1 = \frac{\frac{q_{10}}{q_{01}}}{1+\frac{q_{10}}{q_{01}}}. \tag{9}$$

Since q_{01} represents the probability of unbinding the substrate S from the BS during unit time, we can identy it with k_{-1}. In addition, the probability of binding S to the BS is proportional to the concentration $[S]$ of the substrate with k_1 being its rate coefficient. Thus,

$$q_{01} = k_{-1}, \quad q_{10} = k_1 [S], \tag{10}$$

which gives

$$p_0 = \frac{K_1}{K_1 + [S]}, \quad p_1 = \frac{[S]}{K_1 + [S]}, \quad K_1 = \frac{k_{-1}}{k_1}. \tag{11}$$

Since p_1 denotes the probability of the state of the enzyme with S being bound (ES), the total number of ES molecules is given by $p_1 [E]_T$, where $[E]_T$ denotes the concentration of the total enzyme, which is assumed to be constant. Therefore, since the maximum reaction rate is given by $k_2 [E]_T$, the ratio of the reaction rate υ to its maximum $\upsilon_{\max}$ is given by

$$\frac{\upsilon}{\upsilon_{\max}} = \frac{k_2 p_1 [E]_T}{k_2 [E]_T} = p_1 = \frac{[S]}{K_1 + [S]}, \tag{12}$$

which is the celebrated *Michaelis-Menten equation*. Here, the rapid equilibrium assumption used in deriving the Michaelis-Menten equation [40] has been replaced with the notion of rapid convergence of the probability distribution to a stationary one. Note that we consider the behavior of a single enzyme molecule rather than the collective behavior of the enzyme and enzyme-substrate complex. In this way, a stochastic formulation of biological regulation gives an alternative interpretation of the Michaelis-Menten equation. It should be noted that we have not assumed the enzyme concentration is constant, which is required to derive Michaelis-Menten equation in the traditional way.

It is not difficult to see that the standard deviation $\sigma_\upsilon = \sqrt{\bar{\upsilon}^2 - \bar{\upsilon}^2}$ is given by

$$\sigma_\upsilon = k_2 \frac{\sqrt{K_1 [S]}}{K_1 + [S]},$$

which gives a rough estimate of the precision of the approximation (12). This is a bonus of the stochastic version of the Michaelis-Menten equation that we have derived.

Example 2. (Allosteric enzyme with an activator [40])
An allosteric enzyme e with one binding site for an activator A in addition to the one for a substrate S is described as

$$\begin{array}{ccc} E + S & \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} & ES \overset{k_p}{\to} E + P \\ + & & + \\ A & & A \\ {\scriptstyle k_{-2}\uparrow\downarrow k_2} & & {\scriptstyle k_{-3}\uparrow\downarrow k_3} \\ EA + S & \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} & ESA \overset{k_{pA}}{\to} EA + P \end{array} \tag{13}$$

The enzyme has now the two binding sites, one for the substrate and the other for the activator A. The enzyme has four states $S_0 = (\phi\phi)$, $S_1 = (S\phi)$, $S_2 = (SA)$, $S_3 = (\phi A)$. The state transition diagram is shown in Fig. 1(b). We can associate the transition probabilities with the rate constants in (13) as

$$\begin{aligned} &q_{10} = k_1 [S],\ q_{01} = k_{-1},\ q_{21} = k_3 [A],\ q_{12} = k_{-3}, \\ &q_{32} = k_{-4},\ q_{23} = k_4 [S],\ q_{03} = k_{-2},\ q_{30} = k_2 [A], \end{aligned} \tag{14}$$

where $[S]$ and $[A]$ denote the concentrations of the substrate and the activator, respectively. The transition matrix is given by

$$Q = \begin{bmatrix} D_0 & q_{01} & 0 & q_{03} \\ q_{10} & D_1 & q_{12} & 0 \\ 0 & q_{21} & D_2 & q_{23} \\ q_{30} & 0 & q_{32} & D_3 \end{bmatrix}. \tag{15}$$

where $D_0 = -(q_{10} + q_{30})$, $D_1 = -(q_{01} + q_{21})$, $D_2 = -(q_{12} + q_{32})$, $D_3 = -(q_{03} + q_{23})$. The rate of the target chemical reactions producing P is given by

$$\gamma = k_p p_1 + k_{pA} p_2$$

The actual computation of the stationary probability distribution for this example is done in Example 6, where you see the result is very complex.

Example 3. (Allosteric enzyme with both activator and inhibitor)
The reaction scheme is written as

$$\begin{array}{ccccc} EI + S & \underset{k_{-7}}{\overset{k_7}{\rightleftarrows}} & ESI & \overset{k_{pI}}{\to} & EI + P \\ {\scriptstyle k_6\uparrow\downarrow k_{-6}} & & {\scriptstyle k_5\uparrow\downarrow k_{-5}} & & \\ I & & I & & \\ + & & + & & \\ E + S & \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} & ES & \overset{k_p}{\to} & E + P \\ + & & + & & \\ A & & A & & \\ {\scriptstyle k_{-2}\uparrow\downarrow k_2} & & {\scriptstyle k_{-3}\uparrow\downarrow k_3} & & \\ EA + S & \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} & ESA & \overset{k_{pA}}{\to} & EA + P \end{array} \tag{16}$$

where I denotes the inhibitor. The enzyme has six states, namely, $S_0 = (\phi\phi)$, $S_1 = (\phi A)$, $S_2 = (SA)$, $S_3 = (S\phi)$, $S_4 = (SI)$, $S_5 = (\phi I)$. The state transition diagram is shown in Fig. 1(c). The right part is identical to the diagram of Fig. 1(a). The state transition matrix is given by

$$Q = \begin{bmatrix} D_0 & q_{01} & 0 & q_{03} & 0 & q_{05} \\ q_{10} & D_1 & q_{12} & 0 & 0 & 0 \\ 0 & q_{21} & D_2 & q_{23} & 0 & 0 \\ q_{30} & 0 & q_{32} & D_3 & q_{34} & 0 \\ 0 & 0 & 0 & q_{43} & D_4 & q_{45} \\ q_{50} & 0 & 0 & 0 & q_{54} & D_5 \end{bmatrix}, \tag{17}$$

where $D_0 = -(q_{10} + q_{30} + q_{50})$, $D_1 = -(q_{01} + q_{21})$, $D_2 = -(q_{12} + q_{32})$, $D_3 = -(q_{03} + q_{23} + q_{43})$, $D_4 = -(q_{34} + q_{54})$, $D_5 = -(q_{05} + q_{45})$. The rate of the target chemical reactions is given by

$$\gamma = k_{pA} p_2 + k_p p_3 + k_{pI} p_4$$

The computation of the stationary probability distribution is done in Example 6.

Example 4. (*Lac* operon)
The *lac* operon, which has been a subject of genetic molecular biology for more than fifty years ([5][38][48][52]), can be regarded as a biological regulator. A number of different models have been proposed for it including a very complicated model with seven *cis*-regulatory elements [38]. The standard model has three binding sites [52]. One is the promoter with RNAP as its unique binding factor.

There are two binding factors associated with the *lac* operon, cAMP-CRP and LacI, which have their own binding sites. They are independent of each other. cAMP-CRP is an activator of the operon, while LacI is a repressor. Binding of LacI at its site prevents the other factors from binding to their sites.

Denote the RNAP, cAMP-CRP and LacI by U_1, U_2 and U_3, respectively. Then, the *lac* operon has the following five states;

$$S_0 = (\phi\phi\phi),\ S_1 = (U_1\phi\phi),\ S_2 = (U_1U_2\phi),$$
$$S_3 = (\phi U_2\phi),\ S_4 = (\phi\phi U_3).$$

The transition diagram is shown in Fig. 2. The transcription rate is given by

$$\gamma = \gamma_1 p_1 + \gamma_2 p_2,$$

where γ_1 and γ_2 are the transcription rates corresponding to S_1 and S_2.

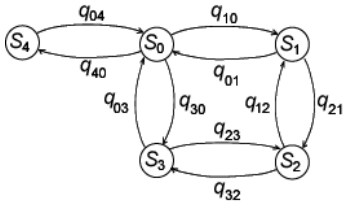

Figure 2: State transition diagram of *lac* operon.

Recently, Santillan et al. proposed a lac operon model with five binding sites [38]. The number of binding patterns of its *cis*-regulatory units is at least 50. Figure 3 shows the transition diagram which is very complicated. We need an efficient computational theory to handle complex operons like this.

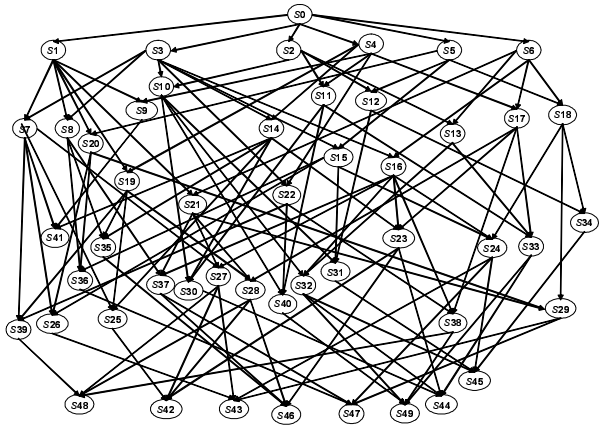

Figure 3: Transition diagram of *lac* operon with 50 states.

3 Conservation of probability flows in stationary distribution and loop-free transition diagram

In this section, we introduce the notion of probability flow and discuss its meaning in solving the SSE (4). The SSE (4) can be written as

$$\sum_{j \in \mathbf{A}_i} (q_{ij} p_j - q_{ji} p_i) = 0, \quad i = 0, 1, \cdots, N. \tag{18}$$

The term $q_{ij}p_j - q_{ji}p_i$ represents the net probability of the state transition from $S_j \in \mathbf{A}_i$ to S_i. Hence, it is reasonable to call it the *probability flow from* S_j *to* S_i, which is denoted by

$$\rho_{ij} = q_{ij} p_j - q_{ji} p_i. \tag{19}$$

If $\rho_{ij} > 0$, the probability flows from S_j to S_i and if $\rho_{ij} < 0$, it flows oppositely. Clearly, it is *skew-symmetric*, i.e.,

$$\rho_{ij} + \rho_{ji} = 0. \tag{20}$$

The SSE (18) is represented in terms of probability flow as

$$\sum_{j \in \mathbf{A}_i} \rho_{ij} = 0, \quad \forall i, \tag{21}$$

which implies that the probability flows are conserved at each state node. In other words, the net probability flow incoming to S_i (the sum of positive ρ_{ij}) is equal to the outgoing flow from S_i (the sum of negative ρ_{ij}), as is illustrated in Figure 4. Here,

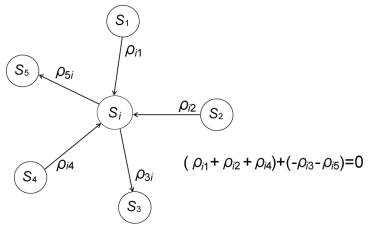

Figure 4: Conservation of probability flows.

it is important to notice that the probability flow is directed.

In chemical kinetics, one usually assumes that the equilibrium state satisfies the relations

$$q_{ij}p_j = q_{ji}p_i, \quad \forall j \in \mathbf{A}_i, \quad \forall i, \tag{22}$$

if we interpret q_{ij} as the kinetic rate coefficient of the reaction $S_j \to S_i$. The relations (22) are called *detailed balance* in chemical kinetics [19]. Due to (19), the detailed balance holds if and only if $\rho_{ij} = 0$, $\forall i$, $j \in \mathbf{A}_i$; i.e., all the probability flows vanish. The relation (22) is written as

$$p_i = r_{ij}p_j \tag{23}$$

where r_{ij} is called the *transition ratio* from S_j to S_i and is defined as

$$r_{ij} = \frac{q_{ij}}{q_{ji}}. \tag{24}$$

The transition ratio (TR) is directed as shown in Fig. 5(a) and corresponds to the equilibrium coefficient of the chemical reaction $S_j \leftrightarrows S_i$. It is sometimes convenient to describe the transition diagram in terms of the transition ratios (TR), instead of transition probabilities, as is shown in Fig. 5(b). We call such a diagram a *TR diagram* to distinguish it from the usual transition diagram. Note that

$$r_{ij} = r_{ji}^{-1}. \tag{25}$$

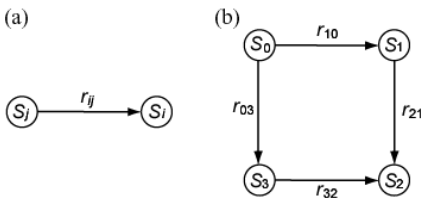

Figure 5: TR Diagrams. (a) Description of TR, (b) TR diagram of Fig. 1(b).

The direction of an edge can be freely assigned in TR diagrams, but it must be consistent with the direction of the TR, in the sense that an edge with r_{ij} as its TR must be directed from S_j to S_i. Now, we shall prove that the probability flows vanish at all edges unless the transition diagram has a loop. Here, a loop is defined in the context of the TR diagram. In other words, if the transition diagram does not have a loop (loop-free), all the probability flows vanish. To see this remarkable fact, assume that there exists an edge e with a non-zero probability flow and a state S_j is connected to this edge. Then, due to the conservation of probability flow at S_j, there exists another edge connecting S_j to another state S_i with a non-zero probability flow. The same reasoning for S_i leads one to conclude that S_i must be connected to a new state S_k by an edge with a non-zero probability flow (Fig. 6). Repeating the procedure creates a sequence of states $S_j \to S_i \to S_k \to \cdots$, which are connected by edges with non-zero probability flows. Since the number of states is finite, the sequence must visit a state which has already appeared in the sequence. Hence, the graph must contain a loop.

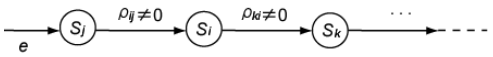

Figure 6: Sequence of edges with non-zero probability flow.

For a while, we concentrate on the loop-free case where all the probability flows vanish. Take two arbitrary states S_i and S_j. If there are two different paths connecting S_i and S_j, it means that there is a loop composed of a path from S_j to S_i and one from S_i to S_j. Therefore, if the transition diagram is loop-free, the state S_i is connected to S_j through a unique path. Let this path be composed of l state: $S_j = S_{i_1} \to S_{i_2} \to \cdots \to S_{i_l} = S_i$. We can define the transition probability from S_j to S_i as

$$\bar{q}_{ij} = q_{ii_{l-1}}q_{i_{l-1}i_{l-2}} \cdots q_{i_2 j}, \tag{26}$$

where the overbar is used to distinguish the transition probability computed along a path from those between adjacent nodes. In this way, we can compute the transition probability for any state pair S_i and S_j which are not necessarily adjacent to each other.

We can extend the detailed balance (22) to any pair of states by using the extended transition probabilities (26).

To see this, assume that three states S_0, S_1, and S_2 are connected by a path, i.e., $S_1 \in \mathbf{A}_0$ and $S_2 \in \mathbf{A}_1$. Then, the detailed balance (22) implies $q_{10}p_0 = q_{01}p_1$ and $q_{21}p_1 = q_{12}p_2$. Multiplying each side yields $q_{10}p_0q_{21}p_1 = q_{01}p_1q_{12}p_2$. Cancelling p_1 from both sides gives $q_{21}q_{10}p_0 = q_{01}q_{12}p_2$, or equivalently, $\bar{q}_{20}p_0 = \bar{q}_{02}p_2$. This relation suggests that the detailed balance holds in terms of the generalized transition probability (26) even for state pairs which are not adjacent to each other. It is not difficult to see that this is indeed the case by repeatedly applying it through the unique path connecting the two states, i.e.,

$$\bar{q}_{ij}p_j = \bar{q}_{ji}p_i, \quad \forall i, j. \tag{27}$$

The above relation includes the detailed balance (22) as a special case. Therefore, we call the relation (27) the *generalized detailed balance*. Taking $j = 0$ in (27), we can represent p_j as

$$p_i = \bar{r}_{i0}p_0, \tag{28}$$

where $\bar{r}_{ij}$ is defined as a generalization of (24) with q_{ij} being replaced by $\bar{q}_{ij}$ given by (26), i.e.,

$$\bar{r}_{ij} = \frac{\bar{q}_{ij}}{\bar{q}_{ji}} = r_{ii_{l-1}}r_{i_{l-1}i_{l-2}}\cdots r_{i_2 j} \tag{29}$$

Here, the overbar is again used to distinguish it from r_{ij} given in (24) for state pairs adjacent to each other. $\bar{r}_{ij}$ is also called the transition ratio (TR) from S_j to S_i. Note that the relation (25) is extended to

$$\bar{r}_{ij} = \bar{r}_{ji}^{-1}.$$

From the normalization constraint (5) and (28), the statitonary probability distribution is simply given by

$$p_i = \frac{\bar{r}_{i0}}{\sum_{j=0}^{N}\bar{r}_{j0}}, \quad i = 0, 1, \cdots, N \tag{30}$$

where $\bar{r}_{00} = 1$.

We sum up the discussion in this section as follows:
If the transition diagram is loop-free, the following facts hold:
(1) the probability flows vanish at every edge.
(2) The stationary distribution is given by (30).
(3) The generalized detailed balance (27) holds for the stationary distribution.

Example 5. (Loop-free Transition Diagram)
Consider a transition diagram of Fig. 7, which is loop-free. According to (28), we have $p_1 = q_{10}p_0$, $p_2 = r_{21}r_{10}p_0$, $p_3 = r_{31}r_{10}p_0$, $p_4 = r_{43}r_{31}p_0$, $p_5 = r_{50}p_0$ with

$$p_0 = \frac{1}{1 + r_{10}(1 + r_{21} + r_{31} + r_{43}r_{31}) + r_{50}}$$

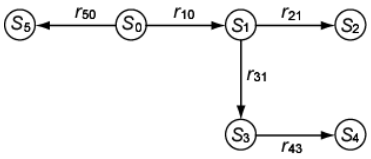

Figure 7: A loop-free transition diagram.

The above reasoning can be applied to a loop-free subgraph of the general (not loop-free) transition diagram. If a loop-free subgraph is attached to a state S_i which belongs to a loop, we can write down the probability of the state of that subgraph according to (30) with p_0 being replaced by p_i. As an example, consider the transition diagram of Fig. 8. The state S_j is a part of a loop which will be discussed in the next section. If the state S_i is not in any loop, then it is a part of a path connecting S_i and the terminal state S_k. Thus,

$$p_i = \bar{r}_{ij}p_j, \tag{31}$$

because the probability flow at any edge on the path connecting S_j and S_k vanishes. This can be shown by directly applying the argument of this section to the subgraph of Fig. 8.

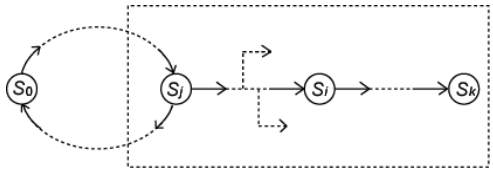

Figure 8: A loop-free subgraph.

4 Wegscheider condition

In the preceding section, we showed that the probability flows vanish if the transition diagram is loop-free. In that case, the stationary distribution is simply calculated from (30). There are cases where the probability flows vanish even if the transition diagram contains loops.

In order to discuss this issue, we consider the case where the whole transition diagram is a loop, as shown in Fig. 9(a). The corresponding TR diagram is shown in Fig. 9(b). The state transition matrix

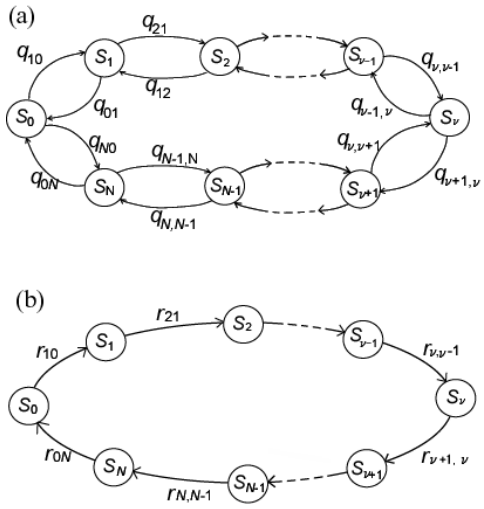

Figure 9: Loop transition Diagram. (a) Transition probability description, (b) transition ratio description.

is given by

$$Q = \begin{bmatrix} D_0 & q_{01} & 0 & \cdots & 0 & q_{0N} \\ q_{10} & D_1 & q_{12} & \cdots & 0 & 0 \\ 0 & q_{21} & D_2 & \cdots & 0 & 0 \\ & \cdots & \cdots & & & \\ q_{N0} & 0 & 0 & \cdots & q_{N,N-1} & D_N \end{bmatrix} \quad (32)$$

with $D_i = -(q_{i+1,i} + q_{i-1,i})$, $i = 1, 2, \cdots, N-1$, $D_0 = -(q_{10} + q_{N0})$, $D_N = -(q_{0N} + q_{N-1,N})$.

Each state has only two edges, one of which receives the probability flow, the other dispatches it. Due to the conservation of probability flows (21), the incoming probability flow and the outgoing one are equal. Hence, each edge of the loop has the same amount of probability flow, which is denoted by ρ. Taking its direction to be clockwise, that is, taking the probability flow from S_{i-1} to S_i to be positive, the probability flow is given by

$$\rho = q_{i,i-1}p_{i-1} - q_{i-1,i}p_i, \quad i = 1, 2, \cdots, N. \quad (33)$$

Thus, we have the following recursion:

$$p_i = r_{i,i-1}p_{i-1} - \frac{\rho}{q_{i-1,i}}, \quad i = 1, 2, \cdots, N, \quad (34)$$
$$p_0 = r_{0N}p_N - \frac{\rho}{q_{N,0}}.$$

In order to obtain a general representation of stationary probabilities, we introduce the numbers $\xi_0, \xi_1, \cdots, \xi_N$ defined sequentially by

$$\xi_0 = 0$$
$$\xi_i = r_{i,i-1}\xi_{i-1} - \frac{1}{q_{i-1,i}}, \quad i = 1, 2, \cdots, N \quad (35)$$

Then, due to (34) and (35), $p_i = r_{i,i-1}p_{i-1} + \rho(\xi_i - r_{i,i-1}\xi_{i-1})$, Therefore, we have

$$p_i - \rho\xi_i = r_{i,i-1}(p_{i-1} - \rho\xi_{i-1}),$$

which implies $p_i - \rho\xi_i = \bar{r}_{i0}p_0$. Here, $\bar{r}_{i0} = r_{i,i-1}r_{i-1,i-2}\cdots r_{10}$. Thus, p_i is represented as

$$p_i = \bar{r}_{i0}p_0 + \rho\xi_i, \quad i = 1, \cdots, N. \quad (36)$$

If $\rho = 0$, the above relations are identical to (28). Thus, equation (36) represents the stationary distribution as a sum of the probability distribution for the case with zero probability flow and a correction term due to non-zero probability flow. Since $\rho = q_{0N}p_N - q_{N0}p_0$, the relation (36) for $i = N$ implies $\rho = q_{0N}(\bar{r}_{N0}p_0 + \rho\xi_N) - q_{N0}p_0 = q_{N0}(r_{0N}\bar{r}_{N0} - 1)p_0 + \rho q_{0N}\xi_N$. Therefore, we get

$$\rho(1 - q_{0N}\xi_N) = -(1 - \eta)q_{N0}p_0 \quad (37)$$

where η is given by

$$\eta = r_{0N}\bar{r}_{N0} = r_{0N}r_{N,N-1}\cdots r_{10} \quad (38)$$

Since $\xi_1 = -1/q_{01} < 0$ and $r_{ij} > 0$, $\xi_N < 0$. Hence, $1 - q_{0N}\xi_N > 0$. Therefore, we conclude that the probability flow vanishes, i.e., $\rho = 0$, if and only if $\eta = 1$.

The number η is the product of all the TRs along the loop. If we use the analogy between the TR and the equilibrium coefficient in chemical kinetics, the condition $\eta = 1$, i.e.,

$$r_{10}r_{21}\cdots r_{0N} = 1, \quad (39)$$

corresponds to the condition derived almost a century ago by Wegscheider. The condition (39) is referred to as the *Wegscheider condition* [22]. We call the product η of (38) the *Wegscheider product* for a loop as in Fig. 9. The Wegscheider product is in some sense directed. The representation (38) defines the Wegscheider product clockwise in the context of Fig. 9. It can also be defined in the opposite direction $\eta' = r_{N0}r_{N-1,N}\cdots r_{01}$. Due to (25), $\eta' = \eta^{-1}$. Since $\eta = 1$ implies $\eta' = 1$, the Wegscheider condition can be represented by computing the product in either clockwise or counter clockwise direction.

The Wegscheider condition is a property of a loop that requires the transition ratio along the total loop to be unity. This condition is equivalently written in terms of transition probabilities as

$$q_{10}q_{21}\cdots q_{0N} = q_{N0}q_{N-1,N}\cdots q_{01}. \quad (40)$$

The left-hand side represents the transition probability from a state to itself along a clockwise contour, while the right-hand side represents that along

a counter clockwise contour. The identity (40) gives a stochastic interpretation of the Wegscheider condition.

Taking the normalization condition (5) into account, we have

$$p_i = \frac{\bar{r}_{i0} + \rho\xi_i}{\sum_{j=0}^{N}\big(\bar{r}_{j0} + \rho\xi_j\big)}, \quad i = 0, 1, \cdots, N, \tag{41}$$

where $r_{00} = 1$, $\xi_0 = 0$. If the Wegscheider condition holds, then $\rho = 0$ and the stationary distribution becomes identical to (30).

The most common method of quantifying the regulatory activities of operons is based on thermal equilibrium theory [3][4], which assigns a Gibbs free energy to each binding pattern of the binding sites. This corresponds to assigning a stationary probability distribution a priori, rather than constructing a Markov process by assigning the transition probabilities between states. The transition ratio may then be defined as the ratio between the state probabilities. In that case, the Wegscheider condition obviously holds. To see this, consider a loop $S_0 \to S_1 \to S_2 \to S_0$. Then, $r_{10} = p_1/p_0$, $r_{21} = p_2/p_1$, $r_{02} = p_0/p_2$. Thus, $r_{10}r_{21}r_{02} = 1$. This proves that the Wegscheider condition is consistent with thermal equilibrium theory. The relation (41), however, suggests more general types of equilibrium.

5 Edge removal and modifier

In this section, we consider the meaning of the numbers ξ_i in (35) and try to interpret (36) in the context of edge removal. From (35), $q_{i-1,i}\xi_i - q_{i,i-1}\xi_{i-1} = -1$ and $q_{i,i+1}\xi_{i+1} - q_{i+1,i}\xi_i = -1$. These relations yield $-q_{i,i-1}\xi_{i-1} - q_{i,i+1}\xi_{i+1} + (q_{i+1,i} + q_{i-1,i})\xi_i = 0$. Since $\xi_0 = 0$ and $\xi_1 = -1/q_{01}$, we have, from (32),

$$Q\xi = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} (1 - q_{0N}\xi_N), \tag{42}$$

where $\xi = \begin{bmatrix}\xi_0 & \xi_1 & \cdots & \xi_N\end{bmatrix}^T$. To further investigate the meaning of the vector ξ, we consider a transition diagram which is obtained from the original diagram of Fig. 9 by eliminating the edge connecting S_0 and S_N (Fig. 10(a)). The modified transition diagram has no loop, and hence, the stationary distribution is given by $\bar{p}_i = \bar{r}_{i0}p_0$, following the discussion in the preceding section. The transition matrix Q' corresponding to the modified diagram of Fig. 10(a) is obtained by taking $q_{0N} = q_{N0} = 0$ in (32). From (42), we have

$$Q'\xi = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{43}$$

We call the vector ξ satisfying (43) with $\xi_0 = 0$ *the modifier* of the loop corresponding to the edge $S_N \to S_0$. It is computed through the recursion formula (35). We shall generalize it in the sequel.

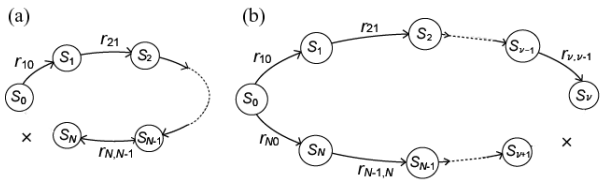

Figure 10: Reduced transition diagrams. (a) Edge $S_N \to S_0$ is eliminated, (b) edge $S_\nu \to S_{\nu+1}$ is eliminated.

If we let p' be the vector whose $(j+1)$th element p_i is given by (28), which represents the stationary probability distribution corresponding to the modified loop-free diagram of Fig. 10(a), equation (36) can be written as

$$p = p' + \rho\xi. \tag{44}$$

This is the general form of the stationary probability distribution. It holds for a loop transition diagram as well as for general transition diagram with multiple loops, as is discussed in the next section. Note that the stationary distribution is represented by a stationary distribution p' corresponding to the modified loop-free diagram with a correcting term $\rho\xi$ due to non-zero probability flow.

Equation (44) is obtained by cutting the edge $S_N \to S_0$ as in Fig. 10(a). Here, we can show that equation (44) is also derived by eliminating any edge $S_\nu \to S_{\nu+1}$ of the loop (Fig. 10(b)). Since $\bar{r}_{i0}$ is defined as the product of each TR along the path connecting S_i to S_0, we have

$$\bar{r}_{i0} = \begin{cases} r_{10}r_{21}\cdots r_{i,i-1} & \text{if} \quad i \le \nu. \\ r_{N0}r_{N-1,N}\cdots r_{i,i-1} & \text{if} \quad i \ge \nu+1. \end{cases} \tag{45}$$

Now, the modifier is defined as

[*Forward Recursion*]

$$\xi_0 = 0$$

$$\xi_i = r_{i,i-1}\xi_{i-1} - \frac{1}{q_{i-1,i}}, \quad i = 1, 2, \cdots, \nu \tag{46}$$

[*Backward Recursion*]

$$\xi_{N+1} = 0. \quad q_{N+1,N} = q_{0N}$$

$$\xi_i = r_{i,i+1}\xi_{i+1} + \frac{1}{q_{i+1,i}}, \quad i = N, N-1, \cdots, \nu+1. \tag{47}$$

The recursion formula (46) computes the modifier elements of the path connecting S_0 to S_ν, while the recursion formula (47) computes those of the path connecting S_0 to $S_{\nu+1}$. The paths are oppositely directed, but both recursions are essentially the same except the signs of the added term. We call (46) *forward recursion*, while (47) *backward recursion*. It is straightforward to see that $\xi = \begin{bmatrix}\xi_0 & \xi_1 & \cdots & \xi_N\end{bmatrix}^T$ satisfies

$$Q\xi = \begin{bmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ -q_{\nu+1,\nu} & q_{\nu,\nu+1} \\ q_{\nu+1,\nu} & -q_{\nu,\nu+1} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_\nu \\ \xi_{\nu+1} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since $q_{\nu+1,\nu} = q_{\nu,\nu+1} = 0$ in the modified transition diagram, we have

$$Q'\xi = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \cdots \nu+1 \\ \cdots \nu+2 \\ \\ \\ \end{matrix} \tag{48}$$

It is not difficult to see that the stationary distribution is given by (36) with $\bar{r}_{i0}$ and ξ_i being represented by (45) and (46)(47), respectively, by applying the arguments of the preceding section to the forward and backward recursions separately.

The modifier is extended to any edge $S_\nu \to S_{\nu+1}$ and can be computed through (46)(47). The stationary probability distribution is given in this case by (44), where p' corresponds to the stationary probability distribution of the reduced loop-free diagram. Since $\rho = 0$ holds under Wegscheider condition, we obtain the following result:

The stationary distribution of a loop is unchanged if an edge is removed, provided that the Wegscheider condition (39) or (40) holds.

In other words, the computation of the stationary distribution for a loop transition diagram is reduced to the loop-free case where a very simple form (30) is already available by eliminating an edge, provided that the Wegsheider condition holds. We shall extend this remarkable property to the general transition diagram with multiple loops in the next section.

Example 6. (Stationary Distribution of Example 2)

This is the case of $N = 3$ in the above algorithm. Remove the edge $e : S_2 \to S_3 (\nu = 2)$. The numbers $\xi_i, i = 0, 1, 2, 3$, are given respectively by forward recursion

$$\xi_0 = 0, \ \xi_1 = -\frac{1}{q_{01}},$$
$$\xi_2 = -\frac{r_{21}}{q_{01}} - \frac{1}{q_{12}} = -\frac{q_{21} + q_{01}}{q_{01}q_{12}} \tag{49}$$

and by backward recursion

$$\xi_3 = \frac{1}{q_{03}}$$

The stationary probability distribution is calculated to be

$$p_1 = r_{10}p_0 - \frac{1}{q_{01}}\rho$$
$$p_2 = \bar{r}_{20}p_0 - \rho\xi_2 = r_{21}r_{10}p_0 - \frac{q_{01} + q_{21}}{q_{01}q_{12}}\rho \tag{50}$$
$$p_3 = r_{30}p_0 + \frac{1}{q_{03}}\rho.$$

The Wegscheider product η is given by

$$\eta = r_{03}r_{32}r_{21}r_{10} = \frac{q_{03}q_{32}q_{21}q_{10}}{q_{30}q_{23}q_{12}q_{01}}.$$

The probability flow $\rho = q_{32}p_2 - q_{23}p_3$ is given by

$$\rho = \frac{q_{03}q_{32}q_{21}q_{10} - q_{30}q_{01}q_{12}q_{23}}{q_{01}q_{12}(q_{03} + q_{32}) + q_{03}q_{23}(q_{01} + q_{21})}p_0$$

6 Eliminating loops from the transition diagram and derivation of the flow equation

Now, we consider the general case where multiple loops exist, and derive a method of computing probability flows. If some edges are removed, the transition diagram becomes loop-free. Although the selection of such loops is not unique, the minimum

number of such edges required to make the transition diagram loop-free is unique. As an example, consider a TR diagram of Fig. 11(a). The elimination of the edges $e_1 : S_2 \to S_3$ and $e_2 : S_4 \to S_5$ makes the diagram loop-free, as is illustrated in Fig. 11(b). We can eliminate, say, the edges $S_0 \to S_1$ and $S_3 \to S_4$ to make the diagram loop-free. The elimination of two edges is necessary and sufficient to make the diagram loop-free in this case.

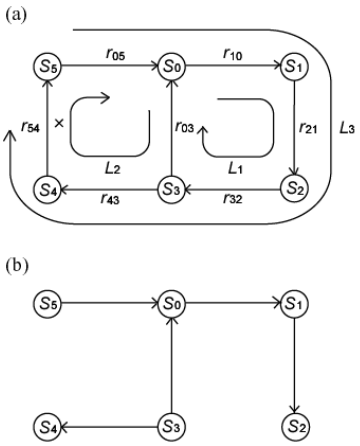

Figure 11: An example with three loops. (a) original diagram, (b) modified diagram.

Now, we recover the eliminated edges in the reduced loop-free diagram one by one. At each recovery step, at least one new loop is recovered and we select one such loop. These loops constitute *basis loops*. In case of Fig. 11, the recovery of edge $e_1 : S_2 \to S_3$ recovers the loop $L_1 : S_0 \to S_1 \to S_2 \to S_3 \to S_0$. Then, recovery of $e_2 : S_4 \to S_5$ recovers $L_2 : S_0 \to S_3 \to S_4 \to S_5 \to S_0$ and $L_3 : S_0 \to S_1 \to S_2 \to S_3 \to S_4 \to S_5 \to S_0$. Thus, we can choose L_1 and L_2 as basis loops. L_1 and L_3 can also be basis loops. In this respect, each edge selected to make the transition diagram loop-free corresponds to one basis loop.

Now, let us consider an algebraic representation for edge removal. Consider an edge $e : S_j \to S_i$. The probability flow of the edge e from S_j to S_i is given by (19), which is alternatively represented as

$$\rho_{ij} = \sigma(i,j)^T p, \tag{51}$$

where $\sigma(i,j)$ is an $(N+1)$-vector given by

$$[\sigma(i,j)]_k = \begin{cases} q_{ij}, & k = j+1 \\ -q_{ji}, & k = i+1 \\ 0, & \text{otherwise.} \end{cases} \tag{52}$$

Note that p_i appears at the $(i+1)$-th place instead of at the i-th place in the vector p because of the existence of p_0 as its first element. Removal of e from the transition diagram amounts to removal of ρ_{ij} and $\rho_{ji} = -\rho_{ij}$ from the $(i+1)$-th and the $(j+1)$-th components of Qp, respectively. Therefore, removal of e corresponds to removal of the matrix

$$\Delta Q := -\delta(i,j)\sigma(i,j)^T, \tag{53}$$

where $\delta(i,j)$ is an $(N+1)$-dimensional vector given by

$$[\delta(i,j)]_k = \begin{cases} -1, & k = i+1 \\ 1, & k = j+1 \\ 0 \end{cases} \tag{54}$$

The reduced transition matrix Q is given by

$$Q' = Q - \Delta Q. \tag{55}$$

Hence, from (53) and (51), we have

$$Q'p = \big(Q + \delta(i,j)\sigma(i,j)^T\big)p = \rho_{ij}\delta(i,j).$$

Assume that the transition diagram contains l basis loops $L_1, L_2, \cdots, L_l$. We can assume that all these loops contain p_0 without loss of generality. Now, choose a set of edges $e_k \in L_k, k = 1, 2, \cdots, l$, such that the elimination of $e_1, e_2 \cdots, e_l$ makes the transition diagram loop-free. We assume that the edge e_k connects the states S_{μ_k} and S_{ν_k} in the direction from S_{μ_k} to S_{ν_k}, i.e., $e_k : S_{\mu_k} \to S_{\nu_k}$. Denote the probability flow associated with e_k by ρ_k, i.e., $\rho_k = \rho_{\nu_k\mu_k}$. Then, from (51), we get

$$\begin{aligned} \rho_k &= q_{\nu_k\mu_k}p_{\mu_k} - q_{\mu_k\nu_k}p_{\nu_k} \\ &= \sigma(\nu_k,\mu_k)^T p. \end{aligned} \tag{56}$$

The transition matrix Q' corresponding to the reduced loop-free transition diagram is thus

$$Q' = Q + \sum_{k=1}^{l} \delta(\nu_k,\mu_k)\sigma(\nu_k,\mu_k)^T. \tag{57}$$

Now we define a vector $\xi(\nu_k,\mu_k)$ satisfying

$$\begin{aligned} &Q'\xi(\nu_k,\mu_k) = \delta(\nu_k,\mu_k), \quad k = 1,2,\cdots,l. \\ &\xi(\nu_k,\mu_k)_0 = 0. \end{aligned} \tag{58}$$

We call $\xi(\nu_k,\mu_k)$ the *modifier* of L_k corresponding to the edge $S_{\mu_k} \to S_{\nu_k}$, which is a generalization of ξ introduced in the preceding section satisfying (43) or (48). Since Q' corresponds to a loop-free transition matirx, its stationary probability distribution p' is given by

$$p'_i = \bar{r}_{i0}p_0, \quad i = 1,2,\cdots,N. \tag{59}$$

where $\bar{r}_{i0}$ denotes the transition ratio from S_0 to S_i taken along a unique path connecting S_0 to S_i in the reduced loop-free transition diagram. From (57) and (58), we get

$$Q'\big(I - \sum_{k=1}^{l} \xi(\nu_k, \mu_k)\sigma(\nu_k, \mu_k)^T\big) = Q.$$

From $Qp = 0$,

$$Q'\big(I - \sum_{k=1}^{l} \xi(\nu_k, \mu_k)\sigma(\nu_k, \mu_k)^T\big)p = 0.$$

From $Q'p' = 0$ and the uniqueness of the stationary distribution, we have

$$\big(I - \sum_{k=1}^{l} \xi(\nu_k, \mu_k)\sigma(\nu_k, \mu_k)^T\big)p = p'.$$

Taking (51) into account, we can now write the stationary probability distribution explicitly as

$$p = p' + \sum_{k=1}^{l} \rho_k \xi(\nu_k, \mu_k), \tag{60}$$

where p' denotes the stationary probability distribution of the reduced loop-free transition diagram given by (59). This generalizes (44). Equation (60) clearly shows the importance of the probability flows reflecting the loop structure of the transition diagram. If the probability flows vanish for each loop, the stationary distribution is identical to that of reduced loop-free diagram given in (30). The modifier $\xi(\nu_k, \mu_k)$ represents the dependence of the stationary distribution on the probability flow.

It remains to compute the probability flow $\rho_1, \rho_2, \cdots, \rho_l$. Premultiplication of (60) by $\sigma(\nu_k, \mu_k)^T$ yields

$$\rho_k - \sum_{m=1}^{l} \sigma(\nu_k, \mu_k)^T \xi(\nu_m, \mu_m)\rho_m = \sigma(\nu_k, \mu_k)^T p'. \tag{61}$$

Now, from the definition of $\sigma(\nu_k, \mu_k)$ (52), we have

$$\begin{aligned}\sigma(\nu_k, \mu_k)^T p' &= q_{\nu_k\mu_k} p'_{\mu_k} - q_{\mu_k\nu_k} p'_{\nu_k} \\ &= q_{\mu_k\nu_k}(-p'_{\nu_k} + r_{\nu_k\mu_k} p'_{\mu_k}) \\ &= -q_{\mu_k\nu_k}\bar{r}_{\nu_k 0}(1 - \bar{r}_{0\nu_k} r_{\nu_k\mu_k}\bar{r}_{\mu_k 0})p_0\end{aligned}$$

As is shown in Fig. 12, the term $\bar{r}_{0\nu_k} r_{\nu_k\mu_k}\bar{r}_{\mu_k 0}$ corresponds to the product of the transition ratios along L_k associated with the edge connecting S_{ν_k} and S_{μ_k}, i.e., the Wegscheider product of L_k. We denote it by η_k. Equation (61) can now be rewritten as

$$\rho_k + \sum_{m=1}^{l} \theta_{km}\rho_m = \alpha_k(1 - \eta_k)p_0, \quad k = 1, 2, \cdots, l, \tag{62}$$

where $\theta_{km} = -\sigma(\nu_k, \mu_k)^T \xi(\nu_m, \mu_m)$, $\alpha_k = -q_{\mu_k\nu_k}\bar{r}_{\nu_k 0}$, and η_k is the Wegscheider product of L_k. Equation (62) implies that the probability flows can be computed by solving only a linear equation of l unknowns, instead of $(N+1)$ unknowns of SSE (3) or (4). We call (62) the *flow equation*.

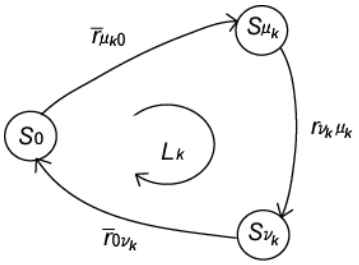

Figure 12: Loop Elimination by Removal of an Edge.

Now, let us summarize the procedure of computing the stationary probability distribution.

Step 1. Find the minimum number of edges $e_k : S_{\mu_k} \to S_{\nu_k}$, $k = 1, 2, \cdots, l$ such that their elimination makes the transition diagram loop-free.

Step 2. Compute the stationary distribution p' via (59) for the reduced loop-free diagram.

Step 3. For each eliminated edge e_k, associate a loop L_k, which is recovered with the edge.

Step 4. For e_k, compute the modifier $\xi(\nu_k, \mu_k)$ that satisfies (58).

Step 5. Solve the flow equation (62) to obtain the probability flows ρ_k, $k = 1, 2, \cdots, l$.

Step 6. Compute p from (60).

Step 7. Compute the exact probability distribution taking the normalization constraint (5) into account.

The computation of the modifier (Step 4) is discussed in Appendix.

Example 7. Consider the transition diagram shown in Fig. 11(a) with 6 states and 3 loops. To eliminate the loops, two edges, $e_1 : S_2 \to S_3$ and $e_2 : S_4 \to S_5$, are removed. The modified transition diagram is shown in Fig. 11(b), which is loop-free.

We choose $L_1 : S_0 \to S_1 \to S_2 \to S_3 \to S_0$, which is created by recovering e_1 in the modified diagram and $L_2 : S_0 \to S_3 \to S_4 \to S_5 \to S_0$ which is created by recovering e_2 in the modified diagram as basis loops. In terms of the notations introduced in this section, $\mu_1 = 2$, $\nu_1 = 3$, $\mu_2 = 4$, $\nu_2 = 5$, $\rho_1 = \rho_{32}$ and $\rho_2 = \rho_{54}$. The transition matrices Q and Q' for the original diagram and the modified diagram are given by

$$Q = \begin{bmatrix} D_0 & q_{01} & 0 & q_{03} & 0 & q_{05} \\ q_{10} & D_1 & q_{12} & 0 & 0 & 0 \\ 0 & q_{21} & D_2 & \boxed{q_{23}} & 0 & 0 \\ q_{30} & 0 & \boxed{q_{32}} & D_3 & q_{34} & 0 \\ 0 & 0 & 0 & q_{43} & D_4 & \boxed{q_{45}} \\ q_{50} & 0 & 0 & 0 & \boxed{q_{54}} & D_5 \end{bmatrix},$$

$$Q' = \begin{bmatrix} D_0 & q_{01} & 0 & q_{03} & 0 & q_{05} \\ q_{10} & D_1 & q_{12} & 0 & 0 & 0 \\ 0 & q_{21} & D_2' & 0 & 0 & 0 \\ q_{30} & 0 & 0 & D_3' & q_{34} & 0 \\ 0 & 0 & 0 & q_{43} & D_4' & 0 \\ q_{50} & 0 & 0 & 0 & 0 & D_5' \end{bmatrix}$$

where $D_0 = -(q_{10} + q_{30} + q_{50})$, $D_1 = -(q_{01} + q_{21})$, $D_2 = -(q_{12} + q_{32})$, $D_3 = -(q_{03} + q_{23} + q_{43})$, $D_4 = -(q_{34} + q_{54})$, $D_5 = -(q_{05} + q_{45})$, $D_2' = -q_{12}$, $D_3' = -(q_{03} + q_{43})$, $D_4' = -q_{34}$, $D_5' = -q_{05}$. The elements of Q encircled by dashed boxes are eliminated in Q'. The modifiers, ξ_1 corresponding to L_1 for e_1 and ξ_2 corresponding to L_2 for e_2, are given as

$$\xi_1 = \begin{bmatrix} 0 \\ 1/q_{01} \\ -r_{21}/q_{01} - 1/q_{12} \\ 1/q_{03} \\ r_{43}/q_{03} \\ 0 \end{bmatrix}, \xi_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1/q_{03} \\ -r_{43}/q_{03} - 1/q_{34} \\ 1/q_{05} \end{bmatrix} \quad (63)$$

The detail of the above calculation is given in Appendix.

According to (62), the flow equation is given by

$$(1 + \theta_{11})\rho_1 + \theta_{12}\rho_2 = \alpha_1(1 - \eta_1)$$
$$\theta_{21}\rho_1 + (1 + \theta_{22})\rho_2 = \alpha_2(1 - \eta_2)$$

where $\theta_{ij} = -\sigma_i^T \xi_j$, $i, j = 1, 2$, with

$$\sigma_1^T = \sigma(3,2)^T = \begin{bmatrix} 0 & 0 & q_{32} & -q_{23} & 0 & 0 \end{bmatrix}$$
$$\sigma_2^T = \sigma(5,4)^T = \begin{bmatrix} 0 & 0 & 0 & q_{54} & -q_{45} & 0 \end{bmatrix}.$$

and

$$\alpha_1 = -q_{23}\bar{r}_{30}, \quad \alpha_2 = -q_{45}\bar{r}_{50}.$$

η_1 and η_2 are Wegscheider products corresponding to L_1 and L_2, respectively, and are given respectively:

$$\eta_1 = r_{10}r_{21}r_{32}r_{03}, \quad \eta_2 = r_{30}r_{43}r_{54}r_{05}.$$

Example 8. (Enzyme with three effectors)
Consider a more complex block transition diagram shown in Fig. 13(a). This transition diagram describes an allosteric enzyme with three different effectors or an operon with three transcription factors. A description like (16) for this example is very complicated and hence, is omitted. Now, we choose four edges $e_1 : S_1 \to S_2$, $e_2 : S_4 \to S_5$, $e_3 : S_6 \to S_7$ and $e_4 : S_7 \to S_8$ to make the diagram loop-free. The reduced diagram is shown in Fig. 13(b). A basis set of loops is composed of the following four loops, $L_1 : S_0 \to S_1 \to S_2 \to S_3 \to S_0$, $L_2 : S_0 \to S_3 \to S_4 \to S_5 \to S_0$, $L_3 : S_0 \to S_1 \to S_6 \to S_7 \to S_0$ and $L_4 : S_0 \to S_7 \to S_8 \to S_5 \to S_0$, as is shown in Fig. 13(a). The stationary probability distribution corresponding to the reduced loop-free diagram Fig. 13(b) is easily calculated according to (59), i.e.,

$$\begin{aligned} &p_1' = r_{10}p_0,\ p_2' = \bar{r}_{20}p_0 = r_{23}r_{30}p_0,\ p_3' = r_{30}p_0, \\ &p_4' = \bar{r}_{40}p_0 = r_{43}r_{30}p_0,\ p_5' = r_{50}p_0, \\ &p_6' = \bar{r}_{60}p_0 = r_{61}r_{10}p_0,\ r_7' = r_{70}p_0, \\ &p_8' = \bar{r}_{80}p_0 = r_{85}r_{50}p_0. \end{aligned}$$

The σ-vectors defined by (52) in this case are $\sigma_1 = \sigma(2,1)$, $\sigma_2 = \sigma(5,4)$, $\sigma_3 = \sigma(7,6)$, $\sigma_4 = \sigma(8,7)$, i.e.,

$$\sigma_1 = \begin{bmatrix} 0 \\ q_{21} \\ -q_{12} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \sigma_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ q_{54} \\ -q_{45} \\ 0 \\ 0 \end{bmatrix}, \ \sigma_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_{76} \\ -q_{67} \\ 0 \end{bmatrix}, \quad \sigma_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_{87} \\ -q_{78} \end{bmatrix}.$$

The modifiers ξ_i, $i = 1, 2, 3, 4$ for L_1, L_2, L_3, L_4 are

computed in Appendix; they are

$$\xi_1 = \begin{bmatrix} 0 \\ -1/q_{01} \\ r_{23}/q_{03} + 1/q_{32} \\ 1/q_{03} \\ r_{43}/q_{03} \\ 0 \\ -r_{61}/q_{01} \\ 0 \\ 0 \end{bmatrix}, \ \xi_2 = \begin{bmatrix} 0 \\ 0 \\ r_{23}/q_{03} \\ -1/q_{03} \\ -r_{43}/q_{03} - 1/q_{34} \\ 1/q_{05} \\ 0 \\ 0 \\ r_{85}/q_{05} \end{bmatrix},$$

$$\xi_3 = \begin{bmatrix} 0 \\ -1/q_{01} \\ 0 \\ 0 \\ 0 \\ 0 \\ -r_{61}/q_{01} - 1/q_{16} \\ 1/q_{07} \\ 0 \end{bmatrix}, \ \xi_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1/q_{05} \\ 0 \\ -1/q_{07} \\ r_{85}/q_{05} + 1/q_{54} \end{bmatrix} \tag{64}$$

The flow equation (62) is given by

$$\begin{bmatrix} 1+\theta_{11} & \theta_{12} & \theta_{13} & 0 \\ \theta_{21} & 1+\theta_{22} & 0 & \theta_{24} \\ \theta_{31} & 0 & 1+\theta_{33} & \theta_{34} \\ 0 & \theta_{42} & \theta_{43} & 1+\theta_{44} \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{bmatrix} = \begin{bmatrix} \alpha_1(1-\eta_1) \\ \alpha_2(1-\eta_2) \\ \alpha_3(1-\eta_3) \\ \alpha_4(1-\eta_4) \end{bmatrix} \tag{65}$$

where $\theta_{ij} = -\sigma_i^T \xi_j$, $i = 1,2,3,4$, $j = 1,2,3,4$, $\alpha_1 = -q_{12}\bar{r}_{20}$, $\alpha_2 = -q_{45}\bar{r}_{50}$, $\alpha_3 = -q_{67}\bar{r}_{70}$, $\alpha_4 = -q_{78}\bar{r}_{80}$, and η_i, $i = 1,2,3,4$, are Wegscheider products for L_i, and are given by $\eta_1 = r_{03}r_{32}r_{21}r_{10}$, $\eta_2 = r_{05}r_{54}r_{43}r_{30}$, $\eta_3 = r_{07}r_{76}r_{61}r_{10}$, and $\eta_4 = r_{05}r_{58}r_{87}r_{70}$. The anti-diagonal elements θ_{14}, θ_{23}, θ_{32} and θ_{41} vanish from the loop structure of Fig. 13.

7 Specification of transition probabilities

In order to obtain a useful representation of the activity of the biological regulator discussed so far, we must give a way to connect transition probabilities q_{ij} with more workable parameters with physical and/or biological meanings. The state transition caused by binding a BF to a BS and that caused by releasing a BF from a BS are essentially different in their nature. The former transition is usually proportional to the concentration of the BF to be

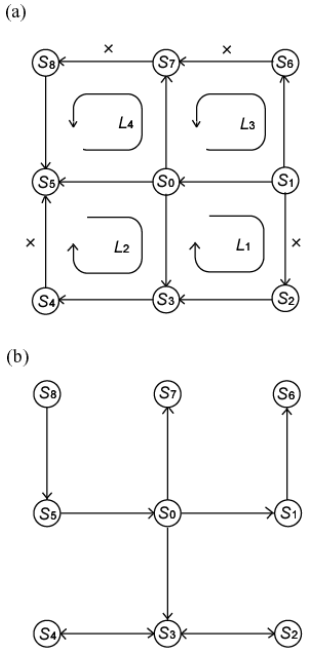

Figure 13: An Example of Transition Diagram. (a) Original Diagram, (b) Reduced Loop-free Diagram. Transition probabilities are omitted.

bound, while the latter does not depend on concentration of BFs but possibly depends on the occupation pattern of other BSs. Therefore, it is natural to assume the following characterization of transition probabilities:

$$q_{ij} = \begin{cases} \alpha_{ij}u_k, & \text{if } S_j \to S_i \text{ is the binding of } U_k, \\ \alpha_{ij}, & \text{if it is a releasing of a BF} \end{cases} \tag{66}$$

for each j and $i \in \mathbf{A}_j$, where α_{ij} denotes a coefficient and u_k the concentration of U_k. The transition probabilities used in Example 1 (equation (10)) and these in Example 2 (equation (14)) are special cases of (66).

From the definition of TR in (24), we have

$$r_{ij} = \begin{cases} \beta_{ij}u_k, & \text{if } S_j \to S_i \text{ is the binding of } U_k, \\ \beta_{ij}u_k^{-1}, & \text{if it is the releasing of } U_k. \end{cases} \tag{67}$$

where $\beta_{ij} = \alpha_{ij}/\alpha_{ji}$. We can derive a concrete form for the overall activity of our biological regulator based on (66), for the case that the Wegscheider condition holds, i.e., the stationary distribution given by (30).

Let S_i be a state where l sites are occupied by U_{i_1}, $U_{i_2}, \cdots, U_{i_l}$ and the remaining $n-l$ sites are empty. Then, according to (28) and (66), the probability p_i is $p_i = a_i u_{i_1} u_{i_2} \cdots u_{i_l} p_0$, where a_i denotes the

product of $\beta_{ij}s$ associated with the binding of U_{i_1}, $U_{i_2}, \cdots, U_{i_l}$. Note that the binding and unbinding of other BFs cannot take place because the path connecting S_0 to S_i does not contain any loop. Now, the normalization constraint (5) yields the overall activity γ defined in (6) as

$$\gamma = \frac{\gamma_0 + \Sigma_{i=1}^{N} \gamma_i a_i u_{i_1} u_{i_2} \cdots u_{i_l}}{1 + \Sigma_{i=1}^{N} a_i u_{i_1} u_{i_2} \cdots u_{i_l}}. \tag{68}$$

This is a familiar form for describing operon regulation, which appears frequently in the literature e.g., [30] [34][40]. For the Lac operon of Fig. 2, the transition probabilities given by (66) implies $q_{10} = \alpha_{10}u_1$, $q_{01} = \alpha_{01}$, $q_{21} = \alpha_{21}u_2$, $q_{12} = \alpha_{12}$, $q_{30} = \alpha_{30}u_2$, $q_{03} = \alpha_{03}$, $q_{40} = \alpha_{40}u_3$ and $q_{04} = \alpha_{04}$. Since only the states bound by U_1 can initiate transcription, $\gamma_0 = \gamma_3 = \gamma_4 = 0$ in (68). Thus, we have

$$\gamma = \frac{\gamma_1 K_1 u_1 + \gamma_2 K_1 K_2 u_1 u_2}{1 + K_1 u_1 + K_1 K_2 u_1 u_2 + K_3 u_2 + K_4 u_3}, \tag{69}$$

where $K_1 = \alpha_{10}/\alpha_{01}$, $K_2 = \alpha_{21}/\alpha_{12}$, $K_3 = \alpha_{30}/\alpha_{03}$ and $K_4 = \alpha_{40}/\alpha_{04}$.

The general form (41), as well as its specification (68), enables us to derive a variety of complicated formulae representing enzymic actions (e.g., [40]) almost immediately. Equation (68) also generalizes the classical MWC model [32] to heterotropic cases.

It is important to notice that the Wegscheider product is constant and does not depend on concentration of any transcription factor given the specification (66). To see this, notice that if the loop contains an edge associated with the binding of U_k, it must contain an edge associated with unbinding of U_k, because the loop must recover the starting state along the path. If the binding transition contains a TR with u_k, then it must contain TR associated with unbinding of U_k which contains u_k^{-1}, so that they cancel out in the Wegscheider product. Thus, we have shown that *the Wegscheider product is always constant* and does not contain the concentrations of BF.

8 Conclusions

The metabolic process is controlled by enzymes whose expressions are controlled by genetic regulations. On the other hand, genetic regulation is controlled by the products of metabolism that determine the cell state. In this sense, the genetic and metabolic regulations are closely linked together to form a huge and complex network of intracellular regulations. It has been desirable to establish a common framework for quantitatively describing these regulations in a unified way. For that purpose, we used the analogy between genetic and metabolic regulations shown in Table 1. The core of the analogy is that the actual computations of the control action are performed through molecular interactions at the regulatory sites between the sites to be bound and the factors to bind or to dissociate.

We formulated a finite Markov process describing both the genetic and metabolic regulations. The most salient feature of this Markov process is its reciprocity in transition probability; that is, for each pair of state, if the transition probability from one state to another is positive, the opposite transition probability is also positive. This property reflects the reversibility of the molecular interactions we are dealing with, and it is the source of many interesting properties of the Markov process we have formulated. The parameters that determine the action of regulations is represented as the average rate of the stationary probability distribution of that Markov process, based on the assumption that the dynamics of molecular interactions that compute the control actions are much faster than the reaction they are regulating. This assumption is analogous to the fast equilibrium assumption used in deriving the classical Michaelis-Menten equation [40]. Actually, our approach can derive a stochastic version of Michaelis-Menten formula which suggests its legitimacy. Moreover, we can derive a quantitative estimate of how precise the Michaelis-Menten equation is by supplying the variance of the stationary distribution.

We introduced a new notion of probability flow associated with each edge of the transition diagram to represent the loop structure of the transition diagram. The state transition diagram is nothing but a representation of the conservation of probability flows at each node. Based on this observation, we derived many interesting properties of the stationary probability distributions.

We proved a simple result that the probability flows vanish if the graph has no loop. This immediately implies that the detailed balance holds for pairs of adjacent states. This was generalized to include pairs of states which are not necessarily adjacent to each other. A very simple graphical method of computing the stationary distribution was derived.

We derived a condition which guarantees the detailed balance even if the transition diagram has a loop. This condition is not new; Wegscheider, a German chemist, discovered this condition a century ago. The condition, which we call the Wegsheider condition in this paper, requires that the prod-

uct of all the transition ratios around a loop be unity. We combined this condition with the probabilistic flow and directly showed that the probability flows vanish under it. This result again gives a probabilistic interpretation of this classical result. The detailed balance has been accepted as an obvious fact in literature of chemical kinetics and enzymology, For instance, a famous book of Segel assumed this condition without even mentioning Wegscheider's name [40]. We gave a stochastic interpretation of this condition.

We derived a simple method of computing the stationary probability distribution for general cases with multiple loops. We showed that if the Wegsheider condition holds for a loop, we can eliminate any of the edges contained in the loop without changing the stationary probability distribution. In other words, if the Wegscheider condition holds for a loop, we can consider a transition diagram with that loop by removing an arbitrary edge within that loop. The stationary probability distribution of the reduced diagram is identical to that of the original diagram.

We derived a purely graphical method of computing the stationary probability distribution based on the notions of probability flow and modifiers. We derived a simple equation for computing the probability flow. Our method dramatically simplifies the classical King and Altman method [24].

Our result suggests that there are two kinds of stationary probability distributions: the one which satisfies the detailed balance, and one which does not. The former class is characterized by zero probability flow, the latter by non-zero probability flows. In the latter case, the probability distribution is fixed, but continuous flows of probability exist in loops, which perhaps reflects a sort of thermal irreversibility of molecular interactions. Exploitation of the bio-chemical characterization of the probability flow is an interesting issue of theoretical biology.

The theoretical framework developed in this paper is based on the observation that the intracellular regulations are embedded in a homogeneous computational medium of molecular interactions. This view is not entirely new, but no serious attempt has been made so far to mathematically formulate it, within the best of our knowledge. A finite Markov process model proposed in this paper captures some essential features of the computational medium and explains the versatility, evolvability and flexibility of the intracellular regulations and various signal transductions. It offers a potential capability to deal with systems in which metabolism and gene expressions are linked and integrated together. We are now exploiting an analytical tool for investigating such systems with greater complexity based on our framework presented in this paper.

References

[1] Ackers, G.K., Johnson, A.D., Shea, M.A., 1982. Quantitative model for gene regulation by λ phage repressor. Proc. Nat. Acad. Sci. 79, 1129-1133.

[2] Beckett, D., 2005. Multilevel regulation of protein-protein interactions in biological circuitry. Phys. Biol. 2, S67-73.

[3] Berg, O.G., von Hippel, P.H., 1987. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. J. Mol. Biol. 193, 723-750.

[4] Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Phillips, R., 2005. Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. 15, 116-24.

[5] Bliss, R.D., Painter, P.R., Marr, A.G., 1982. Role of feedback inhibition in stabilizing the classical operon. J. Theor. Biol. 97, 177-193.

[6] Bolouri, H., Davidson, E.H., 2002. Modeling DNA sequence-based cis-regulatory gene networks. Dev. Biol. 246, 2-13.

[7] Chen, K.C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., Tyson, J.J., 2000. Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol. Biol. Cell 11, 369-91.

[8] Cherry, J.L., Adler, F.R., 2000. How to make a biological switch. J. Theor. Biol. 203, 117-133.

[9] Coppey, M., Benichou, O., Voituriez, R., Moreau, M., 2004. Kinetics of target site localization of a protein on DNA: a stochastic approach. Biophys. J. 87, 1640-1649.

[10] Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436, 588-592.

[11] Endy, D., Brent, R., 2001. Modelling cellular behaviour. Nature 409, 391-395.

[12] Fell, D., 1997. Understanding the Control of Metabolism. Portland Press, London.

[13] Freire E., 2000. Can allosteric regulation be predicted from structure? Proc. Natl. Acad. Sci. USA. 97, 11680-11682.

[14] Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in *Escherichia coli*. Nature 403, 339-342.

[15] Gibson, K.M., Giren, J.A., Head, M.S., 1997. A new class of models for computing receptor-ligand binding affinities. Chem. Biol. 4, 87-92.

[16] Gillespie, D.T., 1977. Stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2341-2361.

[17] Griffith, J.S., 1968. Mathematics of cellular control processes. I. Negative feedback to one gene, II. Positive feedback to one gene. J. Theor. Biol. 20, 202-208, 209-16.

[18] Gunasekaran, K., Ma, B., Nussinov, R., 2004. Is allostery an intrinsic property of all dynamic proteins? Proteins 57, 433-43.

[19] Heinrich, H., Schuster, S., 1996. The regulation of Cellular Systems. Chapman and Hall, New York.

[20] Istrail, S., Davidson, E.H., 2005. Logic functions of the genomic cis-regulatory code. Proc. Natl. Acad. Sci. USA. 102, 4954-4959.

[21] Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genetics 6, 451-464.

[22] Karmarkar, R., Bose, I., 2004. Graded and binary responses in stochastic gene expressions. Phys. Biol. 1, 197-204.

[23] Killer, A.D., 1995. Model genetic circuit encoding autoregulatory transcription factors. J. Theor. Biol. 172, 169-185.

[24] King, E.L., Altman, C., 1956. A schematic method of deriving the rate laws for enzyme-catalyzed reactions. J. Phys. Chem. 60, 1375-1378.

[25] Kitano, H., 2004. Biological robustness. Nat. Rev. Genet. 5, 826-837.

[26] Lloyd, G., Landini, P., Busby, S., 2001. Activation and repression of transcription initiation in bacteria. Essays Biochem. 37, 17-31.

[27] Mackey, M.C., Santillan, M., Yildirim, N., 2004. Modeling operon dynamics: the tryptophan and lactose operons as paradigms. C.R. Biol. 327, 211-24.

[28] McAdams, H.H., Arkin, A., 1997. Stochastic mechanisms in gene expression. Proc. Nat. Acad. Sci. 94, 814-819.

[29] Ming, D., Wall, M.E., 2005. Quantifying allosteric effects in proteins. Proteins 59, 697-707.

[30] Mochizuki, A., 2005. An analytical study of the number of steady states in gene regulatory networks. J. Theor. Biol. 236, 291-310.

[31] Monod, J., 1972. Chance and Necessity, Paperback.

[32] Monod, J., Wyman, J., Changeux, J.P., 1965. On the nature of allosteric transitions: a plausible model. J. Mol. Biol. 12, 88-118.

[33] Munsky, B., Khammash, M., 2006. The finite state projection algorithm for the solution of the chemical master equation. J. Chem. Phys. 124, 044104

[34] Ozbudak, E.M., Thattal, M., Lim, H.M., Shralman, B.I., von Oudenaarden, A., 2004. Multistability in the lactose utilization network of Escherichia coli. Nature 427, 737-740

[35] Pan, H., Lee, J.C., Hilser, V.J., 2000. Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. Proc. Natl. Acad. Sci. USA 97, 12020-12025.

[36] Perutz, M.F., 1989. Mechanisms of cooperativity and allosteric regulation in proteins. Q. Rev. Biophys. 22, 139-237.

[37] Ptashne, M., 1992. A Genetic Switch, Cell & Blackwell Scientific, Cambridge, MA.

[38] Santillan, M., Mackey, M., 2004. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. Biophys. J. 86, 75-84.

[39] Savageau, M.A., 1985. A theory of alternative designs for biochemical control systems. Biomed. Biochim. Acta. 44, 875-880.

[40] Segel, I.H., 1993. Enzyme Kinetics: Behaviour and analysis of rapid equilibrium and steady-state enzyme systems, John Wiley and Sons, INC., New York.

[41] Smolen, P., Boxter, D.A., Byrne, J.H., 2000. Modeling transcriptional control in gene networks–methods, recent results, and future directions. Bull. Math. Biol. 62, 247-292.

[42] Tanaka, R.J., Okano, H., Kimura, H., 2006. Mathematical description of gene regulatory units. Biophys. J. 91, 1235-1247.

[43] Thattai M., van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. USA 98, 8614-8619.

[44] Thomas, R., Thieffry, D. Kauffman, M., 1995. Dynamical behaviour of biological regulatory network – I. Biological role of feedback loops and practical use of the concept of the loop-characteristics state. Bull. Math. Biol. 57, 247-276.

[45] Tyson, J.J., 1978. The dynamics of feedback control circuits in biochemical pathways. Prog. Theol. Biol. 5, 1-61.

[46] van Kampen, N.G., 1992. Stochastic Process in Physics and Chemistry. Elsevier Science, Amsterdam.

[47] Wolf, D.M., Eeckman, F.H., 1998. On the relationship between genomic regulatory element organization and gene regulatory dynamics. J. Theor. Biol. 195, 167-186.

[48] Wong, P., Gladney, S., Keasling, J.D., 1997. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. Biotechnol. Prog. 13, 132-143.

[49] Wyman, J., 1968. Linked functions and reciprocal effects in hemoglobin: A second look. Adv. Protein Chem. 19, 223-286.

[50] Yanofsky, C., 1992. Transcriptional regulation: Elegance in design and discovery. Transcription Regulation I. Cold Spring Harbor Lab. Press, pp.3-24.

[51] Yi, T.M., Huang, Y., Simon, M.I., Doyle, J., 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc. Natl. Acad. Sci. USA. 97, 4649-4653.

[52] Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., Alon, U., 2004 Just-in-time transcription program in metabolic pathways. Nat. Genet.36, 486-491.

Appendix

Computation of the modifier can be done through the simple recursions presented in Section 4. The modifier corresponding to the loop L_k is an $(N+1)$-dimensional vector $\xi_k = \xi(\nu_k, \mu_k)$ satisfying (58). For the case that the transition diagram itself is a loop, we have seen that the recursions formulae (46) and (47) give a vector ξ that satisfies (48), which corresponds to the special case $\mu_k = \nu$, $\nu_k = \nu+1$. The computation of the modifier for a general loop is essentially reduced to (46)(47). To avoid notational complication, we assume that L_k is the loop $S_0 \to S_1 \to S_2 \to \cdots \to S_\pi \to S_0$. Since $e_k : S_{\mu_k} \to S_{\nu_k}$ is in the loop, we assume $\mu_k = \nu$, $\nu_k = \nu + 1$, $\nu \leq \pi - 1$. Then, the components of the modifier corresponding to the states inside the loop L_k are given by (46) and (47). Thus, we have a procedure for computing the components of the modifier corresponding to the states inside the loop. The components of ξ_k outside the loop L_k are computed as follows: let S_j be a state outside L_k. Since the reduced diagram is loop-free, there exists a unique path connecting S_j to S_0. If the path does not have a common state with L_k except S_0, let the component of ξ_k corresponding to S_j be zero. If the path has some states in common with L_k, there is a state S_{k_j} nearest S_j in L_k. Then, the component of ξ_k corresponding to S_j is given by

$$(\xi_k)_j = \bar{r}_{jk_j}\xi_{k_j} \tag{A1}$$

where $\bar{r}_{jk_j}$ is the TR from the state S_{k_j} to S_j and ξ_{k_j} is the component of the modifier corresponding to S_{k_j} which has already been computed. The justification of (A1) is based on the remark at the end of Section 3 (equation (31)).

Example A1. We compute ξ_1 and ξ_2 in Fig. 11. For L_1, the forward recursion corresponding to (46) is given by

$$\begin{aligned}
&\xi_{10} = 0 \\
&\xi_{11} = -\frac{1}{q_{01}} \\
&\xi_{12} = r_{21}\xi_{11} - \frac{1}{q_{12}} = -r_{21}\frac{1}{q_{01}} - \frac{1}{q_{12}}
\end{aligned}$$

and the backward recursion corresponding to (47) is given by

$$\xi_{13} = \frac{1}{q_{03}}.$$

From the graph of Fig. 11(b), the path connecting S_4 to S_0 meets L_1 at S_3, while the path connecting

S_5 to S_0 directly reaches S_0. Hence, we have

$$\xi_{14} = r_{43}\xi_{13} = \frac{r_{43}}{q_{03}}, \quad \xi_{15} = 0.$$

For L_2, the forward recursion is given by

$$\begin{aligned}
&\xi_{20} = 0 \\
&\xi_{23} = -\frac{1}{q_{03}} \\
&\xi_{24} = r_{43}\xi_{23} - \frac{1}{q_{34}} = -\frac{r_{43}}{q_{03}} - \frac{1}{q_{34}}
\end{aligned}$$

and the backward recursion is given by

$$\xi_{25} = \frac{1}{q_{05}}.$$

For the states outside L_2, S_2 and S_1 are directly connected to S_0 without meeting L_2. Hence, $\xi_{22} = \xi_{21} = 0$. Thus, (63) has been confirmed.

Example A2. We compute ξ_1, ξ_2, ξ_3 and ξ_4 corresponding to the loops L_1, L_2, L_3 and L_4 of Example 8.

For L_1, the forward recursion ($S_0 \to S_1$) gives

$$\xi_{11} = -1/q_{01},$$

and the backward recursion ($S_0 \to S_3 \to S_2$) gives

$$\begin{aligned}
\xi_{13} &= 1/q_{03}, \\
\xi_{12} &= r_{23}\xi_{13} + 1/q_{32}.
\end{aligned}$$

The probabilities of the states outside L_1 are given by

$$\begin{aligned}
\xi_{14} &= r_{43}\xi_{13}, \\
\xi_{1i} &= \bar{r}_{i0}\xi_{10} = 0,\ i = 5, 6, 7, 8.
\end{aligned}$$

For L_2, the forward recursion ($S_0 \to S_3 \to S_4$) gives

$$\begin{aligned}
\xi_{23} &= -1/q_{03}, \\
\xi_{24} &= r_{43}\xi_{23} - 1/q_{34},
\end{aligned}$$

and the backward recursion ($S_0 \to S_5$) gives

$$\xi_{25} = 1/q_{05}.$$

The probabilities outside L_2 are given by

$$\begin{aligned}
\xi_{22} &= r_{23}\xi_{23}, \\
\xi_{28} &= r_{85}\xi_{25}, \\
\xi_{2i} &= \bar{r}_{i0}\xi_{20} = 0,\ i = 1, 6, 7.
\end{aligned}$$

For L_3, the forward recursion ($S_0 \to S_1 \to S_6$) gives

$$\begin{aligned}
\xi_{31} &= -1/q_{01}, \\
\xi_{32} &= r_{41}\xi_{31} - 1/q_{14},
\end{aligned}$$

and the backward recursion ($S_0 \to S_7$) gives

$$\xi_{37} = 1/q_{07}$$

. The probabilities outside L_3 is given by

$$\xi_{3i} = \bar{r}_{i0}\xi_{20} = 0,\ i = 2, 3, 4, 5, 8$$

.

For L_4, the forward recursion ($S_0 \to S_7$) gives

$$\xi_{47} = -1/q_{07}$$

, and the backward recursion ($S_0 \to S_5 \to S_8$) gives

$$\begin{aligned}
\xi_{45} &= 1/q_{05}, \\
\xi_{48} &= r_{85}\xi_{45} + 1/q_{54}.
\end{aligned}$$

The probabilities outside L_4 is given by

$$\xi_{4i} = \bar{r}_{i0}\xi_{40} = 0,\ i = 1, 2, 3, 4, 6.$$

The above scheme verifies (64).